



Title	Mining Complex Data Streams
Author(s)	中村, 航大
Citation	大阪大学, 2025, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/101760">https://hdl.handle.net/11094/101760</a>
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">https://www.library.osaka-u.ac.jp/thesis/#closed</a> 大阪大学の博士論文について

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

## 論文内容の要旨

氏名 ( 中村 航大 )	
論文題名	Mining Complex Data Streams (複合データストリームマイニング)

## 論文内容の要旨

Real-world data is scaling to ever larger sizes and faster transaction rates. Flows of data points that are continuously generated with no end in sight and arrive at fast rates-known as data streams-are everywhere, from large-scale online systems to high-frequency sensor networks. With the recent advances in web-scale database management systems and Internet of Things (IoT) applications, concepts and algorithms for data streams are required to address a wider range of real-world scenarios. This raises questions such as: How can we detect abnormal heartbeats with an electrocardiogram in real time? Are there any hidden patterns, rules, and anomalies in a huge collection of shopping logs? Can we predict when failures will occur by analyzing real time sensor data?

In this thesis, we extend the concept of data streams and refer to the data mentioned in the questions above as “complex data streams”. In particular, we discuss and categorize complex data streams from several points of view, namely, (a) sequential data stream, (b) tensor stream, (c) quasi tensor stream, and (d) predictive tensor stream, for which we propose new models, algorithms, and applications.

The first proposed method, TS-SKETCH, focuses on sequential data streams and how to capture multi-scale statistics for real time anomaly detection. Second, we discuss data streams with multiple attributes and their high dimensionality by introducing tensor streams. The proposed method, CUBESCOPE, is a fundamental approach for data compression, pattern discovery, and anomaly detection over tensor streams. Third, we extend the concept of tensor streams to quasi tensor streams in order to handle both categorical and continuous attributes, for which we propose CYBERCSCOPE. The method computes hybrid infinite and finite dimensional decomposition for anomaly detection and pattern discovery. The fourth method, PRISM, focuses on predictive tensor streams and predicts the occurrence of potential events based on the joint optimization of pattern discovery and event prediction.

Throughout extensive experiments using real-world datasets, we conduct both quantitative and qualitative analyses of the four approaches. The results show that the proposed methods can achieve high performance in various data mining tasks compared with state-of-the-art baselines while ensuring scalability for large-scale data streams at high rates. The final chapter discusses future directions and key components in the proposed methods that can open new possibilities for broader data and tasks.

## 論文審査の結果の要旨及び担当者

氏 名 ( 中 村 航 大 )	
	(職) 氏 名
論文審査担当者	主査 教授 櫻井 保志 副査 教授 浦西 友樹 副査 准教授 谷口 一徹 副査 准教授 佐々木 勇和

## 論文審査の結果の要旨

各データ点が継続的かつ高速に取得され、データ長が半無限長となる性質を持つデータは、データストリームと呼ばれ、多くの実世界のシステムにおいて収集・解析されてきた。しかしながら、時間的依存性や高次元性といったような複雑な性質を併せ持つデータストリームに対する解析は依然として困難である。本論文は、そのような複合的なデータストリームに対してパターン検出、異常検知、および、将来イベント予測を行うためのアルゴリズムの開発に取り組んだものである。本論文では、実データを用いた評価実験を実施し、以下の成果を得ている。

## [時系列データストリームを対象とした高速異常検知手法の開発]

IoTセンサデータをはじめとした時系列データストリームの解析では、高速に取得され続けるデータを効率的に解析し、そこで発生するデータの異常な振る舞いをリアルタイムに検出する必要がある。本研究では、ベルヌーイ螺旋関数を基底とした時系列変換に基づき異常区間を検出するアルゴリズムを開発した。提案手法は既存の時系列異常検知手法と比較して、大幅な速度向上を実現しながら高精度に異常を検出することを明らかにした。

## [テンソルストリームのためのパターン検出・異常検知手法の開発]

本研究では、購買ログや人口流動データなどの高次元テンソルストリームに対して、確率的生成モデルに基づくオンラインテンソル分解を応用して、多角的な観点に基づく時系列パターンの発見や突如発生する異常な部分テンソルの検出を行うアルゴリズムを開発した。提案手法は、モデリング、パターン検出、異常検知の観点で既存手法よりも高精度であり、高速に処理を行うことを明らかにした。

さらに、サイバーセキュリティ領域で収集されるデータストリームは、質的データと量的データなど、異なる種類の属性情報が混在するテンソルストリームとなる。そのようなデータに対して、本研究では、確率的生成モデルと情報理論に基づく符号化スキームを応用し、時系列パターン検出と異常検知（ここでは特に、サイバー攻撃検知）を実現した。実験では、提案手法が、実際のサイバー攻撃を検出しながらそれらの攻撃の種類を時系列パターンとして同定できることを明らかにした。

## [テンソルストリームのための将来イベント予測手法の開発]

本研究では、工場機器センサデータなどのテンソルストリームを解析し、機器の故障などの将来発生しうるイベントの発生時刻を予測するアルゴリズムを開発した。提案手法は、確率微分方程式とガウシアングラフィカルモデルを応用し、時系列パターン検出とイベント予測モデルの学習を同時にすることで、データの傾向の変化に適応しながら高精度かつ高速にイベント発生時刻を予測することが可能であることを明らかにした。

以上の研究成果は、大規模かつ多様なデータストリームにおける効率的かつ効果的な解析を可能にし、それらの解析を通して高精度なパターン検出・異常検知・将来イベント予測を実現することを示した。これらのことから、本論文は、データストリーム解析の性能向上、適用範囲の拡大に大きく貢献するものであると期待できる。よって、博士（情報科学）の学位論文として価値のあるものと認める。